



# ANALISIS SENTIMEN DAN PENGGALIAN OPINI



Asrumi  
Didik Suharijadi  
Agustina Dewi Setiari  
Diah Putri Wulanda

# ANALISIS SENTIMEN DAN PENGGALIAN OPINI

Analisis sentimen adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi masyarakat terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atributnya. Analisis sentimen biasanya dilakukan dalam konteks data teks, seperti ulasan produk, tweet, posting media sosial, artikel berita, dan lainnya. Tujuan analisis sentimen adalah untuk memahami bagaimana orang merasakan atau berpikir tentang suatu topik, produk, layanan, merek, atau peristiwa. Analisis sentiment berdasarkan penggalan opini ini sudah banyak dikenal dalam dunia perdagangan sejak sebelum tahun 2000-an. Riset bidang ini sangat produktif karena terdapat banyak ragam aplikasi di berbagai bidang, termasuk aplikasi komersial yang mampu menyediakan banyak data, di web, dan munculnya medsos. Analisis sentiment menjadi pusat penelitian media social yang berdampak pada dunia ilmu manajemen, ilmu politik, ilmu ekonomi, dan ilmu-ilmu social.

Dalam analisis sentimen, teks atau data dianalisis untuk menentukan apakah mengandung sentimen positif, negatif, atau netral. Sentimen ini dapat berkaitan dengan berbagai aspek, termasuk kepuasan pelanggan terhadap produk, reaksi masyarakat terhadap berita, atau persepsi konsumen terhadap merek.

Dalam analisis sentimen, ilmu semantik dalam unsur pemahaman makna kata, makna kalimat, hubungan makna, dan pemahaman konteks turut berperan dalam mengurai teks opini, apakah menunjukkan sentimen positif, negatif atau netral. Dalam analisis sentimen tersebut dapat dilakukan melalui analisis: (1) Preprosesing Teks, yakni membersihkan dan mengolah teks untuk menghilangkan karakter khusus, tanda baca, dan kata-kata yang tidak relevan; (2) Pembobotan Kata, yakni menghitung bobot kata-kata dalam teks untuk menilai sejauh mana kata-kata tersebut mengindikasikan sentimen positif atau negative; dan (3) Analisis Emosi, yakni mengidentifikasi emosi yang terkandung dalam teks, seperti kebahagiaan, marah, sedih, atau takut.

Analisis sentiment dapat berbasis aspek, berbasis leksikon, berbasis perbandingan pendapat, dan berbasis penggalan opini. Buku ini berbicara tentang penggunaan kata-kata, frase, dan kalimat-kalimat dalam dunia ekonomi perdagangan (online) yang disoroti dari segi makna. Hal ini penting untuk mengetahui animo pembeli lewat komentar di media online sebagai parameter perusahaan atau penjual untuk melakukan evaluasi diri. Apakah animo suatu produk itu mendapat sentiment positif, negatif, atau netral.

**ANALISIS SENTIMEN  
DAN PENGGALIAN OPINI**

Asrumi  
Didik Suharijadi  
Agustina Dewi Setiari  
Diah Putri Wulanda



**eureka**  
**media aksara**

**PENERBIT CV.EUREKA MEDIA AKSARA**

## ANALISIS SENTIMEN DAN PENGGALIAN OPINI

**Penulis** : Asrumi  
Didik Suharijadi  
Agustina Dewi Setiari  
Diah Putri Wulanda

**Desain Sampul** : Ardyan Arya Hayuwaskita

**Tata Letak** : Husnun Nur Afifah

**ISBN** : 978-623-151-907-8

Diterbitkan oleh : EUREKA MEDIA AKSARA, NOVEMBER 2023  
ANGGOTA IKAPI JAWA TENGAH  
NO. 225/JTE/2021

**Redaksi:**

Jalan Banjaran, Desa Banjaran RT 20 RW 10 Kecamatan Bojongsari  
Kabupaten Purbalingga Telp. 0858-5343-1992  
Surel : eurekamediaaksara@gmail.com  
Cetakan Pertama : 2023

**All right reserved**

Hak Cipta dilindungi undang-undang  
Dilarang memperbanyak atau memindahkan sebagian atau seluruh isi buku ini dalam bentuk apapun dan dengan cara apapun, termasuk memfotokopi, merekam, atau dengan teknik perekaman lainnya tanpa seizin tertulis dari penerbit.

## ANALISIS SENTIMEN DAN PENGGALIAN OPINI

(Bing Liu. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012, liub@cs.uic.edu)

Buku yang berjudul “Analisis Sentimen dan Penggalian Opini” ini merupakan buku terjemahan dari buku “*Sentiment Analysis and Opinion Mining*” oleh Bing Liu yang diterbitkan oleh Morgan & Claypool Publishers, May 2012. Buku ini dapat digunakan sebagai bahan perkuliahan Semantik Bahasa Indonesia, yang terdiri atas 12 bab.

Buku ini membicarakan (1) Analisis Sentimen sebagai Problem yang Menarik; (2) Masalah Analisis Sentimen; (3) Klasifikasi Sentimen Dokumen; (4) Subjektivitas Kalimat dan Klafifikasi Sentimen; (5) Analisis Sentimen Berbasis Aspek; (6) Generasi eLksikon Sentimen; (7) Rangkuman Opini; (8) Analisis Perbandingan Pendapat; (9) Pencairan dan Pengambilan Opini; (10) Deteksi Spam Opini; (11) Kualitas Ulasan; dan (12) Catatan Penutup. Buku ini berbicara *tentang* penggunaan kata-kata, frase, dan kalimat-kalimat dalam dunia ekonomi perdagangan (*online*) yang disoroti dari segi makna. Hal ini penting untuk mengetahui animo pembeli lewat komentar di media online sebagai parameter perusahaan atau penjual untuk melakukan evaluasi diri. Apakah animo suatu produk itu mendapat sentiment positif, negative, atau netral.

Dengan buku ini, pembelajar Semantik sebagai ilmu linguistik tentang makna-makna tanda bahasa, berupa kata, frase, dan kalimat dapat dikembangkan dalam dunia perdagangan, khususnya tentang animo pembeli atau konsumen dalam mengomentari produk yang telah dibeli. Dengan mengetahui berbagai animo konsumen, perusahaan atau produk dapat mengetahui, membranding, *dan* mengevaluasi serta mengambil langkah-langkah baru yang inovatif agar omset penjualan cenderung positif.



Sudahkah hidup kita bermakna?

“Hidup akan penuh arti dan bermakna, jika bermanfaat bagi orang lain, bukan hanya untuk keluarga dan diri sendiri”

*‘Life will be full of meaning and meaningful, if it is useful for other people, not just for family and yourself’*



## KATA PENGANTAR

Puji syukur dipanjatkan kehadirat Allah SWT., karena atas ridho-Nya penulisan buku “Analisis Sentimen dan Penggalian Opini” ini dapat diselesaikan dengan baik. Tulisan ini merupakan hasil terjemahan dari buku yang berjudul “*Sentiment Analysis and Opinion Mining*” yang ditulis oleh Bing Liu yang diterbitkan oleh Morgan & Claypool Publishers, May 2012. Dalam buku ini terdiri atas 12 bab.

Bab 1 tentang Analisis Sentimen sebagai Problem yang Menarik. Bab 2 tentang Masalah Analisis Sentimen. Bab 3 tentang Klasifikasi Sentimen Dokumen. Bab 4 tentang Subjektivitas Kalimat dan Klasifikasi Sentimen. Bab 5 tentang Analisis Sentimen Berbasis Aspek. Bab 6 tentang Generasi Leksikon Sentimen. Bab 7 tentang Rangkuman Opini. Bab 8 tentang Analisis Perbandingan Pendapat. Bab 9 tentang Pencairan dan Pengambilan Opini Bab 1 tentang Deteksi Spam Opini. Bab 11 tentang Kualitas Ulasan. Bab 12 tentang Catatan Penutup.

Buku ini merupakan hasil terjemahan sebagai bahan analisis makna kata, frase, dan kalimat dalam teks komentar konsumen pada perdagangan *online* untuk mengetahui pandangan mereka terhadap sebuah produk yang telah dibeli. Apakah menunjukkan sentimen positif, negatif, atau netral. Dengan analisis sentiment tersebut, para produsen atau penjual dapat memperbaiki diri untuk meningkatkan omset. Penulis berharap agar buku ini bermanfaat bagi pembelajar mata kuliah Semantik Bahasa Indonesia (khususnya), terlepas dari segala kekurangannya yang masih membutuhkan perbaikan. Pembuatan buku ini dapat membantu mahasiswa untuk mengembangkan diri dalam menganalisis bentuk-bentuk bahasa dalam dunia ilmu yang lain, termasuk dunia politik.

Terlepas dari sejumlah kekurangan yang ada, semoga buku ini banyak memberikan manfaat bagi pembaca. Kritik dan saran sangat dibutuhkan demi kesempurnaan buku ini.

Jember, 5 November 2023

Penulis

DAFTAR ISI

**KATA PENGANTAR..... v**  
**DAFTAR ISI..... vi**  
**DAFTAR GAMBAR..... ix**  
**DAFTAR TABEL .....x**  
**BAB 1 ANALISIS SENTIMEN: PROBLEM YANG MENARIK..1**  
    A. Aplikasi Analisis Sentimen.....2  
    B. Penelitian Analisis Sentimen .....5  
    C. Deteksi Spam Opini..... 11  
    D. Apa yang Akan Terjadi.....12  
**BAB 2 MASALAH ANALISIS SENTIMEN .....14**  
    A. Definisi Masalah.....15  
    B. Rangkuman Opini .....27  
    C. Berbagai Jenis Pendapat.....28  
    D. Subjektivitas dan Emosi.....30  
    E. Sudut Pandang Penulis dan Pembaca .....33  
    F. Ringkasan.....34  
**BAB 3 KLASIFIKASI SENTIMEN DOKUMEN.....35**  
    A. Klasifikasi Sentimen Menggunakan Supervised Learning .....37  
    B. Klasifikasi Sentimen Menggunakan Pembelajaran Tanpa Pengawasan.....43  
    C. Prediksi Peringkat Sentimen .....45  
    D. Klasifikasi Sentimen Lintas Domain.....48  
    E. Klasifikasi Sentimen Lintas Bahasa.....53  
    F. Ringkasan.....56  
**BAB 4 SUBJEKTIVITAS KALIMAT DAN KLASIFIKASI SENTIMEN .....58**  
    A. Klasifikasi Subjektivitas .....60  
    B. Klasifikasi Sentimen Kalimat .....65  
    C. Menangani Kalimat Bersyarat.....68  
    D. Menangani Kalimat Sarkastik .....70  
    E. Subjektivitas dan Klasifikasi Sentimen Lintas Bahasa .71  
    F. Menggunakan Informasi Wacana untuk Klasifikasi Sentimen.....75  
    G. Ringkasan.....76



<b>BAB 5 ANALISIS SENTIMEN BERBASIS ASPEK.....</b>	<b>78</b>
A. Klasifikasi Aspek Sentimen.....	80
B. Aturan Dasar Pendapat dan Semantik Komposisi .....	84
C. Ekstraksi Aspek.....	93
D. Mengidentifikasi Aspek Penggunaan Sumber Daya .	110
E. Perluasan Leksikon Opini dan Ekstraksi Aspek Secara Serentak .....	112
F. Mengelompokkan Aspek ke dalam Kategori.....	116
G. Entitas, Pemegang Opini dan Ekstraksi Waktu .....	120
H. Resolusi Inti dan Disambiguasi Pengertian Kata .....	123
I. Ringkasan .....	126
<b>BAB 6 GENERASI LEKSIKON SENTIMEN .....</b>	<b>129</b>
A. Pendekatan Berbasis Kamus .....	130
B. Pendekatan Berbasis Korpus .....	137
C. Fakta yang Diinginkan dan Tidak Diinginkan.....	143
D. Ringkasan .....	145
<b>BAB 7 RANGKUMAN OPINI.....</b>	<b>147</b>
A. Peringkasan Opini Berbasis Aspek .....	148
B. Penyempurnaan Peringkasan Opini Berbasis Aspek	151
C. Peringkasan Pandangan Kontrasif .....	154
D. Peringkasan Tradisional.....	155
E. Ringkasan .....	156
<b>BAB 8 ANALISIS PERBANDINGAN PENDAPAT .....</b>	<b>158</b>
A. Definisi Masalah.....	159
B. Identifikasi Kalimat Perbandingan .....	163
C. Mengidentifikasi Entitas Pilihan .....	165
D. Ringkasan .....	168
<b>BAB 9 PENCARIAN DAN PENGAMBILAN OPINI .....</b>	<b>170</b>
A. Pencarian Web vs. Pencarian Opini .....	170
B. Teknik Pengambilan Opini yang Ada.....	172
C. Ringkasan .....	176
<b>BAB 10 DETEKSI SPAM OPINI .....</b>	<b>177</b>
A. Jenis Spam dan Spam .....	178
B. Deteksi Spam yang Diawasi.....	184
C. Deteksi Spam Tanpa Pengawasan.....	188
D. Deteksi Spam Kelompok.....	194
E. Ringkasan .....	196

<b>BAB 11 KUALITAS ULASAN.....</b>	<b>198</b>
A. Kualitas sebagai Masalah Regresi .....	199
B. Metode Lainnya.....	200
C. Ringkasan.....	204
<b>BAB 12 CATATAN PENUTUP .....</b>	<b>205</b>
<b>DAFTAR PUSTAKA.....</b>	<b>207</b>
<b>INDEX .....</b>	<b>257</b>
<b>TENTANG PENULIS.....</b>	<b>276</b>



## DAFTAR GAMBAR

Gambar 2.1. Ringkasan Opini Berdasarkan Aspek ..... 27  
Gambar 5.1. Contoh Review Format 1 ..... 95  
Gambar 5.2. Contoh Review Format 2 ..... 96  
Gambar 7.1. Ringkasan Opini Berdasarkan Aspek ..... 150  
Gambar 7.2. Visualisasi Rangkuman Opini Berdasarkan Aspek. 150



## DAFTAR TABEL

Tabel 3.1. Penn Treebank Part-Of-Speech (POS) tags .....	39
Tabel 3.2. Pola Tag POS untuk Mengekstraksi Frasa Dua Kata ...	44
Tabel 5.1. Aturan untuk Ekstraksi Kata Aspek dan Opini.....	114
Tabel 10.1. Ulasan Palsu vs. Kualitas Produk.....	180



# BAB 1 | ANALISIS SENTIMEN: PROBLEM YANG MENARIK

Analisis sentimen, juga disebut makna opini, adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi masyarakat terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atributnya. Ini mewakili ruang permasalahan yang besar. Ada juga banyak nama dan tugas yang sedikit berbeda, misalnya analisis sentimen, penambangan opini, ekstraksi opini, penambangan sentimen, analisis subjektivitas, analisis pengaruh, analisis emosi, penambangan tinjauan, dll. Namun, semuanya kini berada di bawah payung analisis sentimen. atau penambangan opini. Meskipun di industri, istilah analisis sentimen lebih umum digunakan, namun di dunia akademis, analisis sentimen dan penambangan opini sering digunakan. Mereka pada dasarnya mewakili bidang studi yang sama. Istilah analisis sentimen mungkin pertama kali muncul pada (Nasukawa dan Yi, 2003), dan istilah penambangan opini pertama kali muncul pada (Dave, Lawrence dan Pennock, 2003). Namun penelitian mengenai sentimen dan opini muncul lebih awal (Das dan Chen, 2001; Morinaga et al., 2002; Pang, Lee dan Vaithyanathan, 2002; Tong, 2001; Turney, 2002; Wiebe, 2000). Dalam buku ini, kami menggunakan istilah analisis sentimen dan makna opini secara bergantian. Untuk menyederhanakan penyajiannya, dalam buku ini kita akan menggunakan istilah opini untuk menunjukkan opini, sentimen, evaluasi, penilaian, sikap, dan emosi. Namun, konsep-konsep ini tidak setara. Kami akan membedakannya bila

# BAB 3 | KLASIFIKASI SENTIMEN DOKUMEN

Mulai dari bab ini, kami membahas arah atau topik penelitian utama saat ini dan teknik intinya. Klasifikasi sentimen mungkin merupakan topik yang paling banyak dipelajari (lihat juga surveinya (Pang dan Lee, 2008)). Hal ini bertujuan untuk mengklasifikasikan dokumen opini sebagai ekspresi opini atau sentimen positif atau negatif. Tugas ini juga umumnya dikenal sebagai klasifikasi sentimen tingkat dokumen karena menganggap keseluruhan dokumen sebagai unit informasi dasar. Sebagian besar makalah penelitian tentang topik ini mengklasifikasikan ulasan online. Oleh karena itu, kami juga mendefinisikan masalah dalam konteks tinjauan, namun definisi tersebut juga berlaku untuk konteks serupa lainnya.

**Definisi problem:** Dengan adanya dokumen opini yang mengevaluasi suatu entitas, tentukan sentimen keseluruhan pemegang opini tentang entitas tersebut, yaitu menentukan  $s$  yang dinyatakan pada aspek UMUM dalam rangkap lima

$$(\_ \text{ UMUM}, s, \_ \_ ),$$

di mana entitas  $e$ , pemegang opini  $h$ , dan waktu opini  $t$  dianggap diketahui atau tidak relevan (tidak peduli).

Ada dua formulasi berdasarkan jenis nilai yang diambil. Jika  $s$  mengambil nilai kategorikal, misalnya positif dan negatif, maka ini merupakan masalah klasifikasi. Jika  $s$  mengambil nilai numerik atau skor ordinal dalam rentang tertentu, misalnya 1 sampai 5, maka masalahnya menjadi regresi.

ulasan tidak berlabel dari domain sumber dan target, SCL pertama-tama memilih sekumpulan  $m$  fitur yang sering muncul di kedua domain dan juga merupakan prediktor yang baik untuk label sumber (makalah memilih fitur-fitur dengan nilai mutual tertinggi). informasi ke label sumber). Fitur-fitur ini disebut fitur pivot yang mewakili ruang fitur bersama dari dua domain. Kemudian menghitung korelasi setiap fitur pivot dengan fitur non-pivot lainnya di kedua domain. Hal ini menghasilkan matriks korelasi  $W$  dimana baris  $i$  merupakan vektor nilai korelasi fitur non pivot dengan fitur pivot ke- $i$ . Secara intuitif, nilai positif menunjukkan bahwa fitur non-pivot tersebut berkorelasi positif dengan fitur pivot ke- $i$  di domain sumber atau di domain baru. Ini menetapkan korespondensi fitur antara dua domain. Setelah itu, dekomposisi nilai singular (SVD) digunakan untuk menghitung pendekatan linier berdimensi rendah  $\hat{W}$  ( $k$  vektor singular kiri atas, ditransposisikan) dari  $W$ . Kumpulan fitur terakhir untuk pelatihan dan pengujian adalah kumpulan fitur asli  $x$  dikombinasikan dengan  $\hat{W}x$  yang menghasilkan  $k$  fitur bernilai nyata. Pengklasifikasi yang dibuat menggunakan fitur gabungan dan data berlabel di domain sumber harus berfungsi di domain sumber dan target.

Pan dkk. (Pan et al., 2010) mengusulkan metode yang mirip dengan SCL pada tingkat tinggi. Algoritma bekerja dalam pengaturan di mana hanya ada contoh berlabel di domain sumber dan contoh tidak berlabel di domain target. Ini menjembatani kesenjangan antar domain dengan menggunakan algoritma penyesuaian fitur spektral (SFA) untuk menyesuaikan kata-kata khusus domain dari domain yang berbeda ke dalam cluster terpadu, dengan bantuan kata-kata independen domain sebagai jembatan. Kata-kata yang tidak bergantung pada domain seperti kata-kata pivot (Blitzer, Dredze dan Pereira, 2007) dan dapat dipilih dengan cara yang sama. SFA bekerja dengan terlebih dahulu membuat grafik bipartit dengan kata-kata yang tidak bergantung pada domain sebagai satu kumpulan node dan kata-kata khusus domain

# BAB

# 4

## SUBJEKTIVITAS KALIMAT DAN KLASIFIKASI SENTIMEN

Seperti yang dibahas pada bab sebelumnya, klasifikasi sentimen tingkat dokumen mungkin terlalu kasar untuk sebagian besar aplikasi. Kami sekarang beralih ke tingkat kalimat, yaitu mengklasifikasikan sentimen yang diungkapkan dalam setiap kalimat. Namun, tidak ada perbedaan mendasar antara klasifikasi tingkat dokumen dan kalimat karena kalimat hanyalah dokumen pendek. Salah satu asumsi yang sering dibuat peneliti tentang analisis tingkat kalimat adalah bahwa sebuah kalimat biasanya berisi satu opini (walaupun dalam banyak kasus tidak benar). Sebuah dokumen biasanya berisi banyak pendapat. Mari kita mulai diskusi kita dengan contoh ulasan:

*“Saya membeli ponsel Motorola dua minggu lalu. Semuanya baik-baik saja pada awalnya. Suaranya jernih dan masa pakai baterainya lama, meski agak besar. Lalu, kemarin berhenti berfungsi.”*

Kalimat pertama tidak mengungkapkan pendapat karena hanya menyatakan fakta. Semua kalimat lainnya mengungkapkan sentimen eksplisit atau implisit. Catatan tidak ada pendapat yang biasanya dianggap netral.

**Definisi masalah:** Diberikan sebuah kalimat  $x$ , tentukan apakah  $x$  menyatakan pendapat positif, negatif, atau netral (atau tidak sama sekali).



# BAB 5

## ANALISIS SENTIMEN BERBASIS ASPEK

Mengikuti perkembangan alami bab-bab, bab ini harus berfokus pada klasifikasi sentimen tingkat frasa dan kata karena dua bab terakhir membahas tentang klasifikasi tingkat dokumen dan kalimat. Namun, kita serahkan topik itu pada bab berikutnya. Dalam bab ini, kami fokus pada analisis sentimen berbasis aspek karena sekarang saatnya untuk menangani masalah lengkap yang dijelaskan dalam Bab 2 dan banyak sentimen frasa dan kata bergantung pada konteks aspek.

Seperti yang telah kita bahas dalam dua bab sebelumnya, mengklasifikasikan teks opini pada tingkat dokumen atau tingkat kalimat seringkali tidak cukup untuk diterapkan karena tidak mengidentifikasi target opini atau memberikan sentimen pada target tersebut. Sekalipun kita berasumsi bahwa setiap dokumen mengevaluasi satu entitas, dokumen yang memberikan opini positif tentang entitas tersebut tidak berarti bahwa penulis memiliki opini positif tentang seluruh aspek entitas tersebut. Demikian pula, dokumen opini negatif tidak berarti penulisnya negatif terhadap segala hal.

Untuk analisis yang lebih lengkap, kita perlu menemukan aspek-aspeknya dan menentukan apakah sentimennya positif atau negatif pada setiap aspek. Untuk mengekstrak detail tersebut, kita masuk ke tingkat aspek, yang berarti kita memerlukan model lengkap dari Bab 2, yaitu analisis sentimen berbasis aspek (atau penambangan opini), yang juga disebut penambangan opini berbasis fitur di (Hu dan Liu, 2004). Perlu diperhatikan bahwa

Zhu, 2009; Andrzejewski, Zhu dan Craven, 2009; Mukherjee dan Liu, 2012; Zhai et al., 2011). Kami akan membahasnya Bagian 5.6. Namun, menurut saya mereka masih terlalu berpusat pada statistik dan memiliki keterbatasannya sendiri. Hal ini akan bermanfaat jika kita dapat beralih ke bahasa alami dan berpusat pada pengetahuan untuk pendekatan yang lebih seimbang. Arah lainnya adalah mengintegrasikan pemodelan topik dengan beberapa teknik lain untuk mengatasi kekurangannya.

Terlepas dari metode utama yang dibahas di atas dan pada tiga bagian sebelumnya, masih ada pekerjaan lain mengenai ekstraksi aspek. Misalnya, Yi dkk. (2003) menggunakan model bahasa campuran dan rasio kemungkinan untuk mengekstrak aspek produk. Ma dan Wan (2010) menggunakan teori pemusatan dan pembelajaran terawasi. Meng dan Wang (2009) mengekstraksi aspek dari produk spesifikasi, yang merupakan data terstruktur. Kim dan Hovy (2006) menggunakan pelabelan peran semantik. Stoyanov dan Cardie (2008) mengeksploitasi resolusi coreference. Toprak, Jakob dan Gurevych (2010) merancang skema anotasi komprehensif untuk anotasi opini berbasis aspek. Anotasi sebelumnya bersifat parsial dan terutama untuk kebutuhan khusus masing-masing makalah. Carvalho dkk. (2011) membubuhi keterangan pada kumpulan perdebatan politik dengan aspek dan informasi lainnya.

## 5. Memetakan Aspek Implisit

Dalam (Hu dan Liu, 2004), ada dua macam aspek yang diidentifikasi, yaitu aspek eksplisit dan aspek implisit. Namun, hal itu hanya membahas aspek-aspek eksplisit saja. Ingat di Bagian 2.1, kami menyebut aspek yang dinyatakan sebagai kata benda dan frasa kata benda sebagai aspek eksplisit, misalnya, “kualitas gambar” dalam “Kualitas gambar kamera ini bagus.” Semua ekspresi lain yang menunjukkan aspek disebut aspek implisit. Ada banyak jenis ekspresi aspek implisit. Kata sifat dan kata keterangan

# BAB

# 6

# GENERASI LEKSIKON SENTIMEN

Saat ini, sudah cukup jelas bahwa kata dan frasa yang menyampaikan sentimen positif atau negatif berperan penting dalam analisis sentimen. Bab ini membahas cara menyusun daftar kata-kata tersebut. Dalam literatur penelitian, kata sentimen disebut juga kata opini, kata polar, atau kata yang mengandung opini. Kata-kata sentimen positif digunakan untuk menyatakan keadaan atau kualitas yang diinginkan, sedangkan kata-kata sentimen negatif digunakan untuk menyatakan keadaan atau kualitas yang tidak diinginkan. Contoh kata kata sentimen positif adalah indah, menakjubkan, dan menakjubkan. Contoh kata sentimen negatif adalah buruk, buruk, dan buruk. Selain kata-kata individual, ada juga frasa dan idiom sentimen, misalnya merugikan seseorang. Secara kolektif disebut leksikon sentimen (atau leksikon opini). Untuk memudahkan penyajian, mulai sekarang ketika kita mengucapkan kata-kata sentimen, yang kita maksud adalah kata-kata dan frasa individual.

Kata sentimen dapat dibagi menjadi dua jenis, tipe dasar dan tipe komparatif. Semua contoh kata di atas adalah tipe dasar. Kata-kata sentimen yang bertipe komparatif (yang termasuk tipe superlatif) digunakan untuk menyatakan opini komparatif dan superlatif. Contoh kata-kata tersebut adalah lebih baik, lebih buruk, terbaik, terburuk, dll., yang merupakan bentuk komparatif dan superlatif dari kata sifat atau kata keterangan dasarnya, misalnya baik dan buruk. Berbeda dengan kata-kata sentimen tipe dasar, kata-kata sentimen tipe komparatif tidak mengungkapkan

# BAB 7 | RANGKUMAN OPINI

Seperti yang dibahas di Bab 2, dalam sebagian besar aplikasi analisis sentimen, seseorang perlu mempelajari opini dari banyak orang karena karena sifat subjektif dari opini, hanya melihat opini dari satu orang saja biasanya tidak cukup. Diperlukan suatu bentuk ringkasan. Bab 2 menunjukkan bahwa rangkap lima opini memberikan informasi dasar untuk ringkasan opini. Ringkasan seperti ini disebut ringkasan berbasis aspek (atau ringkasan berbasis fitur) dan diusulkan pada (Hu dan Liu, 2004; Liu, Hu dan Cheng, 2005). Sebagian besar penelitian peringkasan opini menggunakan ide-ide terkait. Kerangka kerja ini juga banyak diterapkan di industri. Misalnya, sistem analisis sentimen di Microsoft Bing dan Google Product Search menggunakan bentuk ringkasan ini. Ringkasan keluaran dapat berbentuk terstruktur (lihat Bagian 7.1) atau dalam bentuk tidak terstruktur sebagai dokumen teks pendek.

Secara umum, peringkasan opini dapat dilihat sebagai salah satu bentuk peringkasan teks multi-dokumen. Peringkasan teks telah dipelajari secara ekstensif di NLP (Das, 2007). Namun, ringkasan opini sangat berbeda dengan ringkasan dokumen tunggal atau multi-dokumen tradisional (informasi faktual) karena ringkasan opini sering kali berpusat pada entitas dan aspek serta sentimen mengenai hal tersebut, dan juga memiliki sisi kuantitatif, yang merupakan intinya. ringkasan opini berbasis aspek. Peringkasan dokumen tunggal tradisional menghasilkan teks pendek dari teks panjang dengan mengekstraksi beberapa kalimat

# BAB 8

## ANALISIS PERBANDINGAN PENDAPAT

Selain mengungkapkan pendapat positif atau negatif secara langsung terhadap suatu entitas dan aspek-aspeknya, seseorang juga dapat mengungkapkan pendapat dengan membandingkan entitas serupa. Pendapat seperti ini disebut opini komparatif (Jindal dan Liu, 2006a; Jindal dan Liu, 2006b). Pendapat komparatif berkaitan tetapi juga berbeda dengan pendapat biasa. Mereka tidak hanya mempunyai makna semantik yang berbeda tetapi juga mempunyai bentuk sintaksis yang berbeda. Misalnya, kalimat opini reguler yang umum adalah “Kualitas suara ponsel ini luar biasa,” dan kalimat opini komparatif yang umum adalah “Kualitas suara ponsel Nokia lebih baik daripada iPhone.” Kalimat perbandingan ini tidak mengatakan kualitas suara ponsel mana pun baik atau buruk, tetapi sekadar membandingkannya. Karena perbedaan tersebut, maka pendapat komparatif memerlukan teknik analisis yang berbeda pula. Seperti kalimat biasa, kalimat perbandingan dapat berupa opini atau non opini. Kalimat komparatif di atas beropini karena secara eksplisit mengungkapkan sentimen komparatif penulisnya, sedangkan kalimat “iPhone lebih lebar 1 inci dari ponsel Nokia pada umumnya” tidak mengungkapkan sentimen apa pun. Dalam bab ini, pertama-tama kita akan mendefinisikan masalahnya dan kemudian menyajikan beberapa metode yang ada untuk menyelesaikannya. Kita juga harus mencatat bahwa sebenarnya ada dua jenis opini utama yang didasarkan pada perbandingan: opini komparatif dan opini superlatif. Dalam bahasa Inggris, kata

# BAB 9 | PENCARIAN DAN PENGAMBILAN OPINI

Karena pencarian Web telah terbukti menjadi layanan yang berharga di Web, tidak sulit untuk membayangkan bahwa pencarian opini juga akan sangat bermanfaat. Dua jenis kueri penelusuran opini yang umum adalah:

1. Menemukan opini publik tentang entitas atau aspek tertentu dari entitas tersebut, misalnya, menemukan opini pelanggan tentang kamera digital atau kualitas gambar kamera, dan menemukan opini publik tentang suatu isu politik atau kandidat.
2. Temukan opini seseorang atau organisasi (yaitu, pemegang opini) mengenai entitas tertentu atau aspek entitas (atau topik), misalnya, temukan opini Barack Obama tentang aborsi. Jenis penelusuran ini sangat relevan dengan artikel berita, yang mencantumkan individu atau organisasi yang menyampaikan pendapat secara eksplisit.

Untuk jenis kueri pertama, pengguna cukup memberikan nama entitas atau nama aspek bersama dengan nama entitas. Untuk jenis kueri kedua, pengguna dapat memberikan nama pemegang opini dan nama entitas atau topik.

## A. Pencarian Web vs. Pencarian Opini

Mirip dengan pencarian Web tradisional, pencarian opini juga memiliki dua tugas utama: 1) mengambil dokumen/kalimat yang relevan dengan permintaan pengguna dan 2) memberi peringkat pada dokumen atau kalimat yang

# BAB 10 | DETEKSI SPAM OPINI

Opini dari media sosial semakin banyak digunakan oleh individu dan organisasi untuk membuat keputusan pembelian dan membuat pilihan pada pemilu serta untuk pemasaran dan desain produk. Opini positif sering kali berarti keuntungan dan ketenaran bagi bisnis dan individu, yang sayangnya memberikan insentif yang kuat bagi orang-orang untuk memperlakukan sistem dengan memposting opini atau ulasan palsu untuk mempromosikan atau mendiskreditkan beberapa produk, layanan, organisasi, individu, dan bahkan ide sasaran tanpa mengungkapkan niat mereka yang sebenarnya, atau orang atau organisasi tempat mereka bekerja secara diam-diam. Orang-orang seperti ini disebut spammer opini dan aktivitasnya disebut spamming opini (Jindal dan Liu, 2008; Jindal dan Liu, 2007). Spam opini mengenai isu-isu sosial dan politik bahkan bisa menjadi hal yang menakutkan karena dapat membelokkan opini dan memobilisasi massa ke posisi yang bertentangan dengan adat istiadat hukum atau etika. Dapat dikatakan bahwa seiring dengan semakin banyaknya praktik opini di media sosial, spam opini akan semakin merajalela dan canggih, sehingga menghadirkan tantangan besar dalam pendeteksiannya. Namun, hal-hal tersebut harus dideteksi untuk memastikan bahwa media sosial tetap menjadi sumber opini publik yang terpercaya, dan tidak penuh dengan opini palsu, kebohongan, dan penipuan.

Deteksi spam secara umum telah dipelajari di banyak bidang. Spam web dan spam email adalah dua jenis spam yang paling banyak dipelajari. Namun, spam opini sangat berbeda. Ada

# BAB 11 | KUALITAS ULASAN

Dalam bab ini, kita membahas kualitas ulasan. Topiknya terkait dengan deteksi spam opini, tetapi juga berbeda karena ulasan berkualitas rendah mungkin bukan spam atau ulasan palsu, dan ulasan palsu mungkin tidak dianggap sebagai ulasan berkualitas rendah oleh pembaca karena seperti yang telah kita bahas di bab terakhir, dengan membaca ulasan sangat sulit untuk menemukan ulasan palsu. Oleh karena itu, ulasan palsu juga dapat dianggap bermanfaat atau ulasan berkualitas tinggi jika penipu menulis ulasannya lebih awal dan menyusunnya dengan baik.

Tujuan dari tugas ini adalah untuk menentukan kualitas, kegunaan, atau utilitas dari setiap tinjauan (Ghose dan Ipeirotis, 2007; Kim et al., 2006; Liu et al., 2007; Zhang dan Varadarajan, 2006). Ini adalah tugas yang berarti karena diinginkan untuk memberi peringkat ulasan berdasarkan kualitas atau kegunaan saat menampilkan ulasan kepada pengguna, dengan ulasan yang paling bermanfaat terlebih dahulu. Faktanya, banyak situs agregasi ulasan atau hosting telah mempraktikkan hal ini selama bertahun-tahun. Mereka memperoleh skor kegunaan atau kualitas dari setiap ulasan dengan meminta pembaca memberikan umpan balik yang berguna untuk setiap ulasan. Misalnya, di amazon.com, pembaca dapat menunjukkan apakah ia menganggap suatu ulasan bermanfaat dengan menjawab pertanyaan “Apakah ulasan tersebut bermanfaat bagi Anda?” tepat di bawah setiap ulasan. Hasil masukan dari semua yang merespons kemudian dikumpulkan dan ditampilkan tepat



# BAB 12

## CATATAN PENUTUP

Buku ini memperkenalkan bidang analisis sentimen dan penambangan opini serta mensurvei perkembangan terkini. Karena banyaknya masalah penelitian yang menantang dan beragamnya penerapan praktis, penelitian di lapangan menjadi sangat aktif dalam beberapa tahun terakhir. Hal ini telah menyebar dari ilmu komputer ke ilmu manajemen (Archak, Ghose dan Ipeirotis, 2007; Chen dan Xie, 2008; Das dan Chen, 2007; Dellarocas, Zhang dan Awad, 2007; Ghose, Ipeirotis dan Sundararajan, 2007; Hu, Pavlou dan Zhang, 2006; Park, Lee dan Han, 2007) karena opini tentang produk erat kaitannya dengan keuntungan.

Buku ini pertama kali mendefinisikan masalah analisis sentimen, yang memberikan kerangka umum untuk menyatukan berbagai arah penelitian di lapangan. Kemudian dibahas topik klasifikasi sentimen tingkat dokumen yang dipelajari secara luas, yang bertujuan untuk menentukan apakah suatu dokumen opini (misalnya ulasan) mengungkapkan sentimen positif atau negatif. Hal ini diikuti oleh subjektivitas tingkat kalimat dan klasifikasi sentimen, yang menentukan apakah suatu kalimat mengandung opini, dan jika demikian, apakah kalimat tersebut mengandung opini positif atau negatif. Buku ini kemudian menjelaskan analisis sentimen berbasis aspek yang mengeksplorasi kekuatan penuh dari definisi masalah dan menunjukkan bahwa analisis sentimen adalah masalah multi-segi dengan banyak submasalah yang menantang. Teknik-teknik yang ada untuk menghadapinya

- Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis. Show me the money!: deriving the pricing power of product features by mining consumer reviews. in Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2007). 2007.
- Asher, Nicholas, Farah Benamara, and Yvette Yannick Mathieu. Distilling opinion in discourse: A preliminary study. in Proceedings of the International Conference on Computational Linguistics (COLING-2008): Companion volume: Posters and Demonstrations. 2008.
- Asur, Sitaram and Bernardo A. Huberman. Predicting the future with social media. Arxiv preprint arXiv:1003.5699, 2010.
- Aue, Anthony and Michael Gamon. Customizing sentiment classifiers to new domains: a case study. in Proceedings of Recent Advances in Natural Language Processing (RANLP-2005). 2005.
- Banea, Carmen, Rada Mihalcea, and Janyce Wiebe. Multilingual subjectivity: are more languages better? in Proceedings of the International Conference on Computational Linguistics (COLING-2010). 2010.
- Banea, Carmen, Rada Mihalcea, Janyce Wiebe, and Samer Hassan. Multilingual subjectivity analysis using machine translation. in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008). 2008.
- Bar-Haim, Roy, Elad Dinur, Ronen Feldman, Moshe Fresko, and Guy Goldstein. Identifying and Following Expert Investors in Stock Microblogs. in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2011). 2011.
- Barbosa, Luciano and Junlan Feng. Robust sentiment detection on twitter from biased and noisy data. in Proceedings of the International Conference on Computational Linguistics (COLING-2010). 2010.

- Bautin, Mikhail, Lohit Vijayarenu, and Steven Skiena. International sentiment analysis for news and blogs. in Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM-2008). 2008.
- Becker, Israela and Vered Aharonson. Last but definitely not least: on the role of the last sentence in automatic polarity-classification. in Proceedings of the ACL 2010 Conference Short Papers. 2010.
- Beineke, Philip, Trevor Hastie, Christopher Manning, and Shivakumar Vaithyanathan. An exploration of sentiment summarization. in Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications. 2003.
- Benamara, Farah, Baptiste Chardon, Yannick Mathieu, and Vladimir Popescu. Towards Context-Based Subjectivity Analysis. in Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2011). 2011.
- Bespalov, Dmitriy, Bing Bai, Yanjun Qi, and Ali Shokoufandeh. Sentiment classification based on supervised latent n-gram analysis. in Proceeding of the ACM conference on Information and knowledge management (CIKM2011). 2011.
- Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. Automatic extraction of opinion propositions and their holders. in Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text. 2004.
- Bickerstaffe, A. and I. Zukerman. A hierarchical classifier applied to multi-way sentiment detection. in Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). 2010.
- Bilgic, Mustafa, Galileo Mark Namata, and Lise Getoor. Combining collective classification and link prediction. in Proceedings



- scoring model. in Proceedings of AAAI Conference on Artificial Intelligence (AAAI-2010). 2010.
- Chen, Yubo and Jinhong Xie. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 2008. 54(3): p. 477-491.
- Choi, Yejin, Eric Breck, and Claire Cardie. Joint extraction of entities and relations for opinion recognition. in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006). 2006.
- Choi, Yejin and Claire Cardie. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. in Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009). 2009.
- Choi, Yejin and Claire Cardie. Hierarchical sequential learning for extracting opinions and their attributes. in Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010). 2010.
- Choi, Yejin and Claire Cardie. Learning with compositional semantics as structural inference for subsentential sentiment analysis. in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2008). 2008.
- Choi, Yejin, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. Identifying sources of opinions with conditional random fields and extraction patterns. in Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005). 2005.
- Cilibrasi, Rudi L. and Paul M. B. Vitanyi. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 2007. 19(3): p. 370-383.

- impact of reviews. in Proceedings of the International Conference on Electronic Commerce. 2007.
- Ghose, Anindya and Panagiotis G. Ipeirotis. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 2010.
- Ghose, Anindya, Panagiotis G. Ipeirotis, and Arun Sundararajan. Opinion mining using econometrics: A case study on reputation systems. in Proceedings of the Association for Computational Linguistics (ACL-2007). 2007.
- Gibbs, Raymond W and Herbert L. Colston. Irony in language and thought: A cognitive science reader, 2007: Lawrence Erlbaum.
- Gibbs, Raymond W. On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, 1986. 115(1): p. 3.
- Goldberg, Andrew B. and Xiaojin Zhu. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. in Proceedings of HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing. 2006.
- González-Ibáñez, Roberto, Smaranda Muresan, and Nina Wacholder. Identifying sarcasm in Twitter: a closer look. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:shortpapers (ACL-2011). 2011.
- Greene, Stephan and Philip Resnik. More than words: Syntactic packaging and implicit sentiment. in Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL (NAACL-2009). 2009.
- Griffiths, Thomas L. and Mark Steyvers. Prediction and semantic association. in *Neural information processing systems* 15. 2003.

- Indurkha, Nitin and Fred J. Damerau. Handbook of Natural Language Processing, 2010: Second Edition, Chapman & Hall.
- Jakob, Niklas and Iryna Gurevych. Extracting Opinion Targets in a Singleand Cross-Domain Setting with Conditional Random Fields. in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010). 2010.
- Jia, Lifeng, Clement Yu, and Weiyi Meng. The effect of negation on sentimen analysis and retrieval effectiveness. in Proceeding of the 18th ACM Conference on Information and Knowledge Management (CIKM2009). 2009.
- Jiang, Jay J. and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. in Proceedings of Research in Computational Linguistics. 1997.
- Jiang, Long, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Targetdependent twitter sentimen classification. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL2011). 2011.
- Jijkoun, Valentin, Maarten de Rijke, and Wouter Weerkamp. Generating Focused Topic-Specific Sentimen Lexicons. in Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010). 2010.
- Jin, Wei and Hung Hay Ho. A novel lexicalized HMM-based learning framework for web opinion mining. in Proceedings of International Conference on Machine Learning (ICML-2009). 2009.
- Jindal, Nitin and Bing Liu. Identifying comparative sentences in text documents. in Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR-2006). 2006a.
- Jindal, Nitin and Bing Liu. Mining comparative sentences and relations. in Proceedings of National Conf. on Artificial Intelligence (AAAI-2006). 2006b.

- approaches, 265, 268  
are, 104, 228, 257, 287  
ari, 133, 145, 153  
art, 266  
articles, 272  
as, 234, 235, 245, 276, 283, 288  
aspect, 150, 166, 253, 260, 277,  
281, 284  
aspectevaluation, 253  
aspect-of, 253  
aspects, 263, 291, 293  
aspek implisit, 24, 101, 117,  
118, 119  
aspek sentimen, 121, 122, 123  
assessing, 252  
assignment, 237  
associated, 254  
association, 243, 258, 279, 283  
at, 257, 277, 293  
attitude, 244, 276, 277  
attitude?, 244  
attribute, 241  
attributes, 235  
aturan dasar, 92, 137, 149,  
151, 181  
automatic, 229, 236, 238, 250,  
263, 268  
available, 239
- B**
- bahasa, ii, v, vii, 7, 13, 15, 16,  
23, 42, 45, 52, 57, 58, 59, 60,  
70, 77, 78, 79, 80, 89, 93, 97,  
107, 109, 112, 116, 117, 122,  
126, 129, 133, 138, 139, 144,  
149, 150, 154, 164, 169, 171,  
173, 178, 182, 189, 223  
bahasa alami, 13, 15, 23, 116,  
139, 169, 223  
based, 229, 242, 245, 247, 248,  
265, 267, 271, 278, 279, 285,  
289  
be, viii, 253  
being, 244  
berita, 5, 12, 16, 25, 32, 61,  
133, 168, 184  
better?, 228  
biased, 228  
bilingual, 260, 284  
bin, 17, 21, 40, 49, 61, 198, 216  
biomedical, 239  
bird, 218  
bisnis, 3, 4, 13, 100, 192, 196,  
211, 213, 224  
blenders, 230  
blog, 3, 5, 12, 16, 25, 27, 28,  
39, 46, 61, 131, 132, 160,  
165, 187, 189, 190, 213, 254,  
258, 261, 270, 289, 293  
blogs, 228, 259  
boards, 236  
bollywood, 230  
book, 282  
boom-boxes, 230  
boosting-based, 277  
bootstrapping, 66, 71, 97, 107,  
115, 145, 146, 147, 179  
bottleneck, 238  
bu, ii, v, vi, vii, viii, 1, 5, 11,  
12, 13, 14, 15, 41, 111, 195,  
197, 219, 222, 223  
but, 89, 104, 229  
by, 103, 227, 238, 240, 252,  
265, 267, 270
- C**
- case, 228, 237, 242, 279



- categorization, 242, 243, 271, 277
- challenges, 274
- channel, 233
- character, 287
- characteristics, 241
- characters, 283, 286
- chasing, 240
- chinese, 279
- citation, 270
- class, 261, 271
- classification, 226, 229, 230, 231, 233, 234, 235, 236, 238, 239, 240, 241, 245, 247, 250, 256, 258, 260, 261, 262, 267, 268, 269, 271, 273, 274, 278, 281, 283, 284, 285, 289, 290, 293, 294
- classification?, 238
- classifications, 285, 286
- classifier, 230
- classifiers, 227, 286
- classify, 236
- classifying, 267
- clauses, 89, 287, 288
- closer, 242
- cluster, 26, 54, 55, 111
- clustering, 118, 245, 257
- cognitive, 242
- collection, 250
- collective, 230
- comments, 233, 261
- commerce, 14
- common, 265
- communication, 234, 244, 246
- comparative, 43, 239, 240, 248, 251, 255, 289
- comparing, 258
- composition, 265
- compositional, 235
- comprehensive, 274
- computer, 244, 278
- computer-mediated, 244
- con, 251
- conceptdistance, 265
- conditional, 235, 267
- conference, 229, 234, 241, 243, 245, 260, 264, 270, 273, 281, 285, 293
- confident, 254
- configurations, 250
- congressional, 233
- conjunction, 43
- connotation, 239
- construct, 231
- constructing, 275
- construction, 238, 249, 260
- constructions, 239
- consumer, 227, 234, 255, 272, 291
- content, 278
- contentious, 272
- context, 232, 260, 281
- context-aware, 260
- contextual, 250, 287
- contradictory, 251
- conversation, 274
- conversations, 284
- co-occurrence, 243
- coreference, 117, 134, 135, 136, 268, 278
- corpora, 50, 80, 147, 254, 260, 286
- corpus, 233, 247, 249
- correspondence, 231
- co-training, 267
- create, 265
- criterion, 288

- ekstensif, 77, 79, 86, 118, 130, 159, 182
  - ekstraksi aspek, 92, 101, 102, 108, 110, 117, 118, 119, 121, 126, 137, 138, 169, 186
  - ekstraksi waktu, 130
  - elections, 283
  - element, 234
  - eliminating, 293
  - em, 111
  - email, 4, 78, 192
  - emosi, 1, 4, 5, 15, 32, 34, 35, 36, 46, 64, 147, 154, 157, 158
  - emosional, 5, 34, 35
  - emoticon, 68
  - emotion, 265, 289
  - emotions, 268, 287
  - empirical, 246
  - english, 233
  - ensemble, 284, 289
  - entitas, 1, 7, 8, 11, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 35, 38, 39, 60, 61, 63, 79, 82, 85, 86, 87, 88, 98, 105, 112, 116, 118, 130, 131, 132, 133, 134, 135, 136, 138, 140, 159, 160, 163, 165, 166, 169, 171, 172, 173, 175, 176, 178, 179, 180, 182, 183, 184, 185, 186, 187, 191, 194, 197, 199, 209, 211, 220
  - entitas pilihan, 175, 176, 180
  - entities, 234, 289
  - entity, 255, 256, 272
  - entropi, 80, 165
  - environment, 252, 283
  - estimation, 259
  - evaluasi diri, ii
  - evaluatif, 69
  - evaluation, 265, 266
  - evaluative, 233, 292
  - evoked, 265
  - examples, 281
  - expansion, 250, 256, 272
  - experiment, 249, 254, 285
  - experiments, 235
  - exploit, 293
  - exploration, 229
  - exploring, 234
  - expressed, 252
  - expression, 282
  - expressions, 232, 250, 267, 275, 276, 287
  - extended, 264
  - extension, 279
  - extracted, 281
  - extracting, 235, 245, 252
  - extraction, 226, 229, 234, 235, 236, 241, 253, 261, 265, 269, 275, 276, 287
- ## F
- facets, 262
  - factorization, 264
  - facts, 291
  - fakta, 10, 13, 32, 33, 62, 63, 97, 155, 156, 168, 185, 195, 196, 220
  - fakta yang diinginkan, 32, 33, 97, 155, 156
  - fakta yang tidak diinginkan, 33, 63
  - favorability, 268
  - feature, 104, 240, 243, 259, 261, 271
  - features, 227, 240, 245, 273, 289, 292
  - feedback, 240

graph, 231, 239, 242, 285, 288,  
289  
graph-based, 239, 242,  
285,288, 289  
graphs, 278  
great, 104  
grief, 277  
guidelines, 276

## H

hard, 236  
hashtag, 46, 68, 73, 285  
hashtags, 236  
hedge, 262  
hedges, 240  
helpful, 270, 282  
helpfulness, 241, 252, 259, 264  
hidden, 267, 274  
hierarchical, 230  
highcoverage, 264  
highly, 240  
holder, 287  
holders, 229, 252, 259  
holistic, 237  
hotel, 138, 199, 204, 208, 270  
how, 287  
html, 41, 157  
hybrid, 293  
hyperlink, 192  
hyperlinked, 252

## I

identification, 243, 251, 268,  
279, 288  
identifikasi, 178, 196, 204  
identifikasi kalimat  
perbandingan, 179  
identifying, 244, 257, 291  
ilmu politik, 2

ilmu sosial, 2  
imagination, 270  
impact, 241  
implicit, 242, 283  
imply, 292  
important, 291  
improve, 290  
improved, 238, 261  
in, 11, 101, 226, 227, 228, 229,  
230, 231, 232, 233, 234, 235,  
236, 237, 238, 239, 240, 241,  
242, 243, 244, 245, 246, 247,  
248, 249, 250, 251, 252, 253,  
254, 255, 256, 257, 258, 259,  
260, 261, 262, 263, 264, 265,  
266, 267, 268, 269, 270, 271,  
272, 273, 274, 275, 276, 277,  
278, 279, 280, 281, 282, 283,  
284, 285, 286, 287, 288, 289,  
290, 291, 292, 293, 294  
indexing, 246  
induction, 275  
inference, 235  
influences, 253  
informasi wacana, 81  
information, 105, 238, 243,  
245, 261, 265, 267, 270, 275  
informativeness, 269  
inquirer, 157, 278  
integer, 166, 234, 269  
integration, 260  
intelligence, 233  
intention, 272  
interaksi, 138  
interdependent, 263  
international, 234, 260  
interpretasi, 2  
interpretation, 245  
into, 227

## M

- machine, 40, 228, 230, 231, 238, 268, 271
- machines, 267
- mad, 287
- management, 229, 241, 243, 245, 264, 274, 281, 285
- many, 242
- marked, 265
- market, 232, 236
- marketing, 234
- masalah regresi, 39, 49, 215, 216
- massive, 250, 281
- matrix, 256
- me, vi, 46, 50, 115, 158, 160, 165, 186, 208, 210, 212, 216, 227
- measure, 250
- measures, 265
- mechanical, 265
- media, ii, vi, 2, 3, 4, 12, 25, 102, 131, 138, 192, 193, 197, 213, 227, 252
- media online, ii
- mengidentifikasi, 3, 25, 55, 66, 69, 72, 73, 75, 85, 107, 108, 112, 113, 121, 122, 133, 135, 148, 153, 155, 157, 163, 176, 178, 180, 183, 190, 191, 195, 199, 203, 205, 206, 219
- merchants, 262
- message, 236
- metafora, 2
- method, 238, 239, 241
- methods, 258, 279
- metric, 269
- mincuts, 231
- miner, 264
- minimum, 45, 271
- mining, 76, 103, 161, 183, 227, 237, 238, 241, 243, 252, 255, 258, 271, 294
- minipar, 122, 257
- mix, 234
- mixture, 262
- mo, 226
- model, 4, 11, 25, 46, 51, 56, 60, 72, 73, 85, 87, 109, 110, 111, 112, 113, 114, 115, 117, 130, 131, 133, 136, 143, 166, 167, 178, 189, 202, 203, 205, 212, 215, 217, 219, 234, 245, 246, 249, 257, 259, 263, 264, 266, 273, 280, 281, 289, 292
- model topik, 60, 110, 111, 112, 113, 115
- modeling, 227, 246, 260, 262, 293
- models, 230, 238, 249, 254, 262, 274, 276, 278, 280, 281
- moderating, 272
- modern, 226
- money!, 227
- mood, 232
- more, 96, 98, 228
- most, 104, 282
- motion, 237
- movie, 250, 276
- multi-grain, 281
- multilanguage-comparability, 251
- multilevel, 243
- multi-level, 73
- multilingual, 231, 232, 263
- multiparty, 274
- multiple, 226, 231, 289
- multi-way, 230

- tfidf, 261  
than, 87, 242  
that, 292  
the, 103, 104, 226, 227, 228,  
229, 230, 231, 232, 233, 234,  
235, 236, 237, 238, 239, 240,  
241, 242, 243, 244, 245, 246,  
247, 248, 249, 250, 251, 252,  
253, 254, 255, 256, 257, 258,  
259, 260, 261, 262, 263, 264,  
265, 266, 267, 268, 269, 270,  
271, 272, 273, 274, 275, 276,  
277, 278, 279, 280, 281, 282,  
283, 284, 285, 286, 287, 288,  
289, 290, 291, 292, 293, 294,  
295  
their, 229, 235, 263  
theory, 253, 277  
there, 43, 242  
thesaurus, 231, 265  
third, 270, 276  
thought, 242  
through, 239, 250, 260, 266,  
273, 278  
thumbs, 283  
time, 104  
to, 43, 103, 104, 227, 230, 231,  
232, 233, 236, 237, 240, 244,  
247, 250, 253, 255, 256, 260,  
261, 262, 264, 265, 268, 269,  
270, 271, 274, 278, 279, 283,  
290, 292, 293, 294  
together, 226  
tool, 245  
topic, 227, 230, 258, 260, 278,  
281, 290, 292  
topic-in-set, 227  
topics, 243, 245, 252, 255  
topik, vi, 1, 6, 11, 13, 16, 18,  
19, 20, 37, 40, 41, 55, 59, 60,  
61, 82, 85, 92, 101, 102, 106,  
108, 110, 111, 112, 113, 114,  
115, 116, 128, 129, 130, 153,  
160, 165, 167, 168, 169, 182,  
184, 185, 186, 188, 189, 190,  
191, 220, 222  
track, 261, 270, 289  
tracking, 253, 282  
tradisional, 30, 40, 41, 45, 50,  
53, 68, 73, 110, 118, 159,  
163, 166, 168, 169, 179, 184,  
185, 187, 195, 204, 223  
transcripts, 233  
transfer, 52, 53, 56, 59, 88,  
100, 288, 289  
translation, 228, 238  
trec, 187, 188  
tree, 285  
treebank, 41  
treebased, 267  
trifactorization, 256  
true, 246  
turk, 265  
tutorial, 274  
twitter, 228, 236, 247, 283, 285  
type, 289
- ## U
- ulasan, 3, 5, 6, 10, 12, 15, 17,  
18, 19, 21, 29, 30, 38, 39, 40,  
44, 45, 47, 49, 50, 51, 52, 53,  
58, 59, 61, 62, 63, 68, 75,  
102, 103, 105, 107, 109, 110,  
112, 113, 114, 115, 127, 128,  
131, 132, 138, 150, 160, 162,  
163, 164, 165, 166, 168, 182,  
187, 188, 190, 191, 192, 193,

- 194, 195, 196, 197, 198, 199,  
200, 201, 202, 203, 204, 205,  
206, 207, 208, 209, 210, 211,  
212, 213, 214, 215, 216, 217,  
218, 219, 220, 222
- ulasan palsu, 12, 192, 193,  
194, 195, 196, 197, 198, 199,  
200, 201, 202, 203, 204, 205,  
206, 209, 214, 215, 223
- unannotated, 286
- unification, 249
- unified, 246
- unify, 292
- unlabeled, 231, 260, 269, 294
- unstructured, 264
- unsupervised, 264, 282, 283,  
284
- untagged, 275
- up, 4, 13, 224, 271, 283
- up?, 271
- usage, 103
- use, 103, 121, 286, 289
- usefulness, 241
- user, 77, 197, 198, 199, 200,  
201, 254, 262, 282
- user-generated, 282
- using, 228, 231, 234, 236, 238,  
239, 240, 241, 244, 250, 253,  
257, 259, 265, 267, 268, 269,  
271, 277, 280, 281, 287, 289,  
290
- utterances, 283
- V**
- valence, 250, 273
- validation, 288
- value, 237, 239
- variable, 280
- variables, 267
- vector, 40, 266, 267
- vectors, 240, 261
- Verba, ii
- verbal, 253
- very, 103
- via, 227, 243, 255, 258, 259,  
263, 271, 284, 285
- viability, 283
- view, 286
- views, 251, 272
- volume, 5, 11, 199, 202, 227
- W**
- wacana, 81, 82, 99, 133
- walks, 244
- weak, 287, 288
- weasels, 240
- web, 6, 76, 106, 192, 217, 226,  
234, 236, 248, 251, 258, 260,  
266, 270, 279, 281, 283
- web-derived, 283
- weblogs, 262
- word, 43, 150, 226, 237, 246,  
261, 287, 289
- wordnet, 287
- word-of-mouth, 246
- words, 239, 242, 244, 247, 257,  
265, 268, 279, 280
- work, 226
- workshop, vi, 230
- www, 41, 157, 158, 217, 273
- Y**
- yi, 178

## TENTANG PENULIS



**Dr. Asrumi, M.Hum.** Lahir di Blitar tanggal, 29 Juni 1961 adalah dosen Jurusan Sastra Indonesia Fakultas Ilmu Budaya (semula Fakultas Sastra) Universitas Jember, mulai tahun 1989 sampai sekarang. Lulus sarjana Sastra Indonesia, Fakultas Ilmu Budaya tahun 1986. Mendapat gelar Magister Humaniora (M.Hum.) dari Universitas Sebelas Maret Surakarta tahun 2001 dan gelar Doktor dari Universitas Sebelas Maret Surakarta tahun 2014.

Penulis menekuni bidang linguistik sejak tahun 1989. Matakuliah yang penulis ampu meliputi: Sintaksis Bahasa Indonesia, Morfologi, Semantik Bahasa Indonesia, Dialektologi (S1 & S2), Psikolinguistik, Seminar Bahasa, Linguistik Forensik (S2), dan Bahasa Indonesia. Selain menjalankan kegiatan akademik perkuliahan, penulis menjalankan pengabdian kepada masyarakat, penelitian, dan kegiatan-kegiatan penunjang yang lain.

Penelitian yang penulis lakukan berkaitan dengan bidang morfologi, semantik, sintaksis, dan psikolinguistik. Bidang semantik terkait makna leksikal, gramatikal, konotatif, asosiatif, dan makna kultural. Selain itu, penulis juga sering mengadakan penelitian tentang pembelajaran Bahasa Indonesia, termasuk evaluasi pembelajaran Bahasa Indonesia pada tingkat SD, SMP, dan SMA. Penulis menjadi Tim peneliti Potensi Budaya Osing dan Industri Kreatif dana LPDP tahun 2018-2020. Penulis telah menulis Buku "DERIVASI MORFOLOGI AFIKSASI VERBA DALAM BAHASA OSING; Buku POLA-POLA PEMBENTUKAN KATEGORI VERBA DALAM BAHASA OSING (2020); ALAT-ALAT PENGUNGKAP VERBA RESIPROKAL DALAM BAHASA OSING (Juli 2020); Buku POTENSI BUDAYA OSING DAN INDUSTRI KREATIF dibuat bersama Tim Peneliti LPDP; Bersama Tim Peneliti LPDP juga membuat Modul tentang CERITA RAKYAT; FILM DOKUMENTER; NARASI SEBLANG; PEMBUATAN BATIK; DAN MODUL MANAJEMEN INDUSTRI KREATIF (2019 -2020); PENYAKIT, OBAT, DAN TATACARA PENGOBATAN TRADISIONAL ETNIK OSING-INDONESIA

(DESEMBER 2020); SISTEM MORFOLOGI DAN SINTAKSIS KATEGORI VERBA DEVERBAL DALAM BAHASA OSING (Analisis Pembentukan Kata Berdasarkan Paradigma Model Uhlenbeck) (Desember 2021); VERBA TRANSAKSI: ANALISIS LEKSIKAL DAN SEMANTIK KATA MERAMPOK DAN MENCURI (SARA THORGEN) (2022).

Penulis terlibat dalam kegiatan penunjang, yakni sebagai Ketua Jurusan Sastra Indonesia (tahun 2003-2005), sebagai Pembantu Dekan I Bidang Akademik (tahun 2005-2009), dan sebagai Ketua Laboratorium Sastra Indonesia (2017 - 2018). Penulis aktif dalam kegiatan organisasi bidang keilmuan Masyarakat Linguistik Indonesia (MLI), Himpunan Sarjana Kesusastraan Indonesia (HISKI), dan Komunitas Peneliti Kopi serta Anggota TIM Peneliti LPDP 2017-2019



**Didik Suharijadi, S.S., M.A.** Lahir di Madiun tanggal, 22 Juli 1968 adalah dosen Jurusan Sastra Indonesia Fakultas Ilmu Budaya Universitas Jember, mulai tahun 1998 sampai sekarang. Menyelesaikan sarjana Sastra Indonesia, Fakultas Ilmu Budaya tahun 1995. Memperoleh gelar Magister of Art (M.A.) dari Universitas Gadjah Mada Yogyakarta pada tahun 20016.

Penulis menekuni bidang linguistik sejak tahun 1998. Matakuliah yang penulis ampu meliputi: Pragmatik, Semantik, Semiotika, Sinematografi, dan Bahasa Indonesia. Selain melaksanakan kegiatan akademik perkuliahan, penulis menjalankan pengabdian kepada masyarakat, penelitian, dan kegiatan-kegiatan penunjang yang lain.

Penelitian yang penulis lakukan berkaitan dengan bidang semiotika, semantik, pragmatik, dan Sinematografi. Karya tulis ilmiah yang peneliti buat di antaranya sebagai berikut.

1. HUMOR DALAM SINETRON PARA PENCARI TUHAN (Tesis, bidang Semiotika)
2. KESANTUNAN DALAM ULASAN PRODUK DI MARKETPLACE (Artikel Jurnal, bidang Pragmatik)



3. SINEMATOGRAFI PENGANTAR MENUJU KREATIVITAS AUDIO VISUAL (Buku, bidang Sinematografi)
4. ISTILAH-ISTILAH DALAM REGISTER PENGETAHUAN KELAUTAN NELAYAN MADURA DI DESA PESISIR, KECAMATAN BESUKI, KABUPATEN SITUBONDO (Artikel Jurnal, bidang Semiotika dan budaya)
5. THE EFFECTIVENESS OF TOURISM VLOG COMPETITION ON VISITS TO COCOA ECOTOURISM DESTINATIONS (Artikel jurnal, bidang Sinematografi).

Penulis terlibat dalam kegiatan penunjang, yakni sebagai Kepala Laboratorium Produksi Sastra (2019-2023), Ketua Divisi Publikasi Pusat Studi Gender Universitas Jember (2022-sekarang), Ketua Bidang Sosial Humaniora Kelompok Riset Cocoa Innovation and Development (2017-Sekarang), manajer Jurnal Semiotika (2016-sekarang), editor Journal of Feminism and Gender Study (2020-sekarang).



**Dr. Agustina Dewi S., M.Hum.** Lahir di Madiun tanggal, 18 Agustus 1977 adalah dosen Jurusan Sastra Indonesia Fakultas Ilmu Budaya (semula Fakultas Sastra) Universitas Jember, mulai tahun 2001. Mendapat gelar Sarjana dari Universitas Jember, Magister Humaniora (M.Hum.) tahun 2007 dan gelar Doktor dari Universitas Gajah Mada tahun 2016.

Penulis menekuni bidang linguistik sejak tahun 2002. Matakuliah yang penulis ampu meliputi: Morfologi, Semantik Bahasa Indonesia, Bahasa Indonesia, Etnografi Komunikasi, Bahasa dan Gender, Sosiolinguistik. Selain menjalankan kegiatan akademik perkuliahan, penulis menjalankan pengabdian kepada masyarakat, penelitian, dan kegiatan-kegiatan penunjang yang lain.

Penelitian yang penulis lakukan berkaitan dengan bidang morfologi, semantik, etnografi komunikasi, bahasa dan gender, pragmatik. Penulis menjadi Tim peneliti KeRis (Kelompok Riset) Okara, KeRis Medagen, dan KeRis. Penulis telah menulis buku bersama Tim Riset "Mengenal dan Mencegah Kekerasan Seksual di Lingkungan Perguruan Tinggi (2022); Mamaca: Alih Wahana

Kisah-Kisah Ke Dalam Dua Bahasa (2022); Manuskrip Mamaca Sembilan Naskah Hasil Dokumentasi (2022).



**Diah Putri Wulanda.** Lahir di Probolinggo tanggal, 28 Maret 2001 adalah mahasiswa Jurusan Sastra Indonesia Fakultas Ilmu Budaya Universitas Jember. Sebagai seorang mahasiswa yang tengah menempuh semester ketujuh di jurusan Sastra Indonesia, perjalanan akademik penulis telah sangat terkait dengan eksplorasi dalam bidang kepenulisan, terutama dalam aspek sastra dan linguistik. Selama kuliah, penulis telah menerbitkan beberapa buku dalam berbagai genre, baik fiksi maupun non fiksi. Penulis telah menulis buku yang berjudul: Lelucon Dunia; Karsa Jenggama; Tjantung Sastra; Antologi Puisi “Memikirkan Rindu”; dan Antologi Puisi “Aglocita”.

Penulis aktif berpartisipasi dalam himpunan mahasiswa jurusan yang bertugas untuk mewadahi dan menyalurkan minat bakat mahasiswa di jurusan Sastra Indonesia. Dalam organisasi ini, penulis telah terlibat dalam berbagai kegiatan, seperti workshop kepenulisan, saung pembelajaran dan diskusi dan kegiatan-kegiatan yang membantu mengembangkan kemampuan komunikasi dan kepemimpinan penulis. Dengan pemahaman mendalam dalam bidang linguistik dan minat yang kuat dalam kepenulisan, penulis berkomitmen untuk terus berkembang dalam disiplin ini dan memberikan kontribusi yang berarti dalam dunia sastra dan linguistik.

Selain mengejar studi di perguruan tinggi, penulis aktif merintis dan mengasah kemampuan menulisnya melalui media sosial, khususnya platform Blogspot. Ini adalah wadah di mana penulis dapat mengekspresikan ide, pandangan, dan kisah-kisah yang ingin penulis bagi dengan dunia. Blogging telah menjadi sarana yang sangat bermanfaat bagi penulis untuk mengembangkan keterampilan menulis, memperdalam pemahaman penulis tentang berbagai topik, dan berbagi pengetahuan dengan pembaca.

Penulisan buku ini adalah pengalaman yang sangat istimewa bagi penulis. Ini merupakan kali pertama penulis berkesempatan untuk menerbitkan buku bersama dosen-dosen yang berpengalaman di bidang linguistik. Kerjasama kami dalam mewujudkan buku ini merupakan langkah yang sangat berarti dalam perkembangan karier penulis di dunia kepenulisan. Penulis merasa beruntung bisa belajar dari seseorang yang telah memiliki pengalaman luas dalam bidang ini, dan ini memberikan penulis wawasan yang berharga tentang bagaimana menghadirkan karya dengan kualitas terbaik. Penulis sangat bersemangat untuk terus belajar dan berkembang dalam dunia kepenulisan, serta melanjutkan kolaborasi yang penuh inspirasi ke depannya.

